Scientific Journal of Impact Factor (SJIF): 5.71
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406
International Journal of Advance Engineering and Research Development
Volume 8, Issue 05, May-2021

An Analysis of Resource Utilization in Cloud Computing using Alibaba Cluster Trace

Hemanshi H. Vaniya[1], Kirtikumar J. Sharma[2], Pranay S. Patel[3]
[1]PG Scholar [2,3]Assistant Professor
[1, 2, 3]Computer Engineering Department, Birla Vishvakarma Mahavidyalaya, Vallabh Vidhyanagar

Abstract —*Cloud computing usage is rising day by day due to their 'On Request' service. Increasing the usage of cloud computing, there is a need for better utilization and effective resource allocation. For better utilization, one needs to apply new methods, real environment or setup, or there is a need for real-time cloud datasets. Alibaba group released the dataset 'Alibaba cluster trace' in 2017, which is used in this paper. Alibaba cluster trace dataset consists of machines, online services, and offline services. Dataset has 11101 Container events(online services) and 12935 offline batch jobs located in 1300 machines over 12 hours. All services are completed in 5 minutes interval of 12 hours. The paper 'Imbalance in the cloud: An analysis of alibaba cluster trace' describes the study of dataset in depth. Also, reviewed other various research papers based on alibaba cluster trace dataset which is described in this paper and also surveyed the various methods and algorithms. Assignment of the algorithms is done in this paper for better utilization on 'alibaba cluster trace' dataset.*

Keywords*- alibaba cluster trace-2017, cloud computing, online services, batch jobs, machine utilization, resource allocation*

I. INTRODUCTION

Cloud computing plays an important role in reducing the usage of energy at global range. Cloud computing saves energy and can help the business to significantly lower its carbon footprint, which results as an asset for the better utilization.

The dataset alibaba trace data "ClusterData201708" contains 12 hours period and 1300 machines that run both online and offline services. Here is 11101 online services and 12935 offline jobs. In alibaba trace dataset, there are three types of data which consists of machines, online services and offline services. There are total 6 files in the dataset. Machine is divided into two tables: "machine event" table(Server_event.csv) and "machine utilization" table(Server_usage.csv). Online services are further described in two tables: "service instance event" table(Container_event.csv) and "service instance usage" table(Container_usage.csv). Offline batch jobs are also described in two tables: "instance" table(Batch_instance.csv) and "task" table(Batch_task.csv).

As per the alibaba statements or reports 3 types of works are there. Workload characterizations, New algorithm to assign workload to machines and to CPU cores, Online service and batch jobs scheduler cooperation. 1. Workload characterizations: How can we characterize Alibaba workloads in a way that we can simulate various production workload in a representative way for scheduler studies. 2. New algorithms to assign workload to machines: How we can assign workload to different machines and cpus for better resource utilization and acceptable resource contention. 3. Online services and batch jobs scheduler cooperation: How we can adjust resource allocation between online services and offline services to improve throughput of batch jobs while maintain acceptable resource contention[3].

II. THE DATASET

2.1. Schema of dataset

Figure 1 represents entity relationship diagram of the dataset. Diagram describes the entity, attributes and the relation between entities. Entity represents total files of the dataset that is machine event, machine utilization, service instance event, service instance usage etc. Attribute represents the columns of each files. Relationship represents the relation between machine event and machine usage, service instance event and service instance usage, task table and instance table, this is called one to one relation. Machine usage is related with service instance event and instance table, this is called one to many relation of ER diagram.

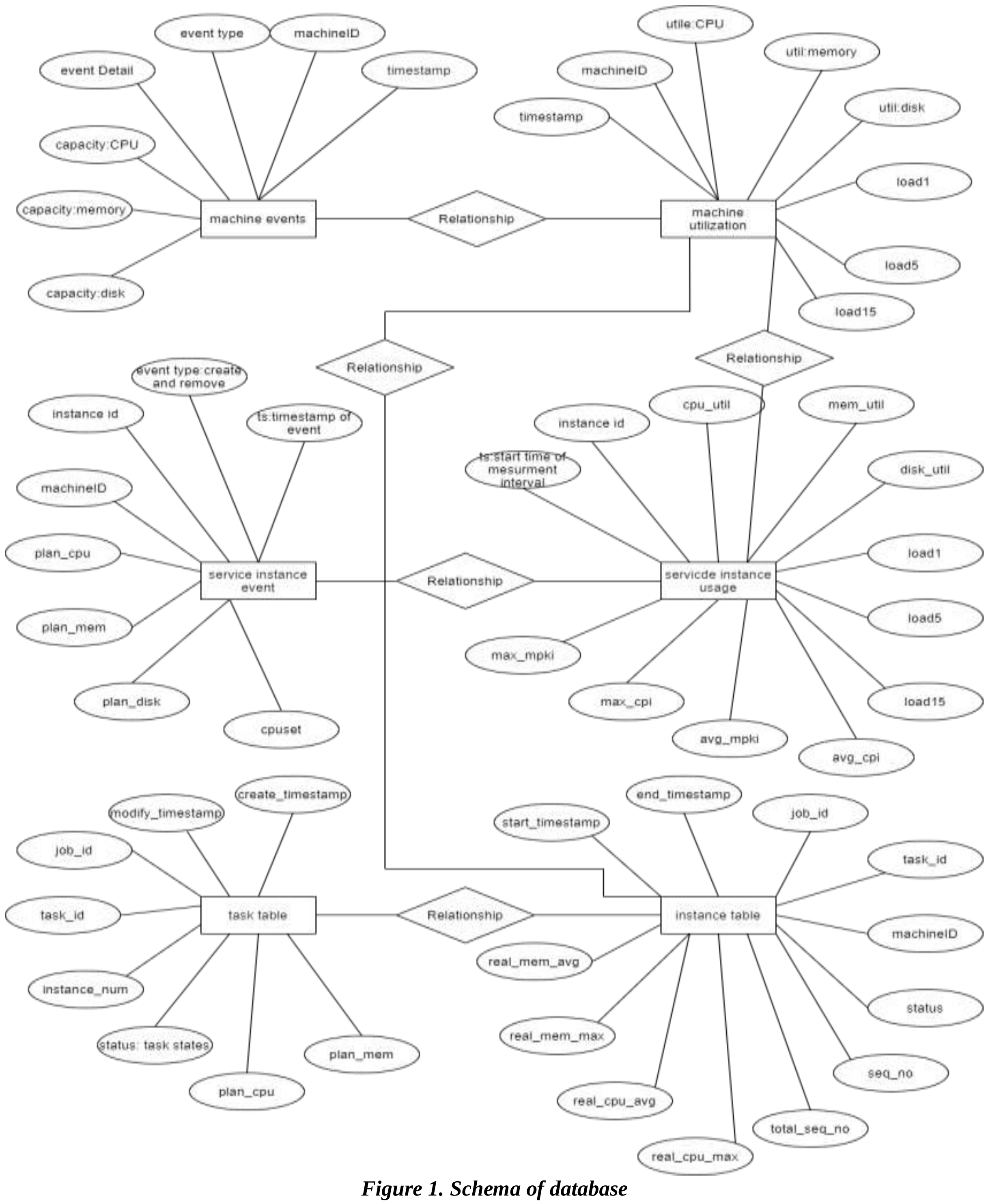


Figure 1. Schema of database

Below describes the dataset in detail

There are three types of data in dataset (1) machine, (2) online services and (3) offline services.

2.2. Machines
There are two files, machine events and machine utilization.

2.2.1 Machine events (server_event.csv)
This trace includes three types of machine events: 1. Add - A machine is available to the cluster. All machines in the trace have an ADD event, and has timestamp of value 0. 2. Softerror - A machine becomes temporarily unavailable . 3. Harderror - A machine becomes unavailable due to hardware failures.
This file has 7 columns as shown in figure 2

In [29]: data

Out[29]:

	timestamp	machineID	event type	event detail	capacity:CPU	capacity:memory	capacity:disk
0	0	1149	add	NaN	64	0.690006	1.0
1	0	1150	add	NaN	64	0.690006	1.0
2	0	1	add	NaN	64	0.689970	1.0
3	0	2	add	NaN	64	0.689970	1.0
4	0	3	add	NaN	64	0.689970	1.0
...	...	...	...	...	...	...	...
1346	74682	1075	softerror	agent_check	0	0.000000	0.0
1347	81179	930	softerror	machine_fail	0	0.000000	0.0
1348	75268	372	softerror	agent_check	0	0.000000	0.0
1349	82706	372	softerror	agent_check	0	0.000000	0.0
1350	82122	1075	softerror	agent_check	0	0.000000	0.0

1351 rows × 7 columns

Figure 2. Dataset summary of machine event file

2.2.2 Machine utilization(server_usage.csv)

Machine utilization has cpu utilization, memory utilization and disk utilization also has load per 1 minute, load per 5 minutes and load per 15 minutes.

In [5]: data

Out[5]:

	timestamp	machineID	util:CPU	util:memory	util:disk	load1	load5	load15
0	39600	265	26.36	29.540000	57.599998	17.46	18.900000	16.70
1	42600	770	49.14	60.099999	41.860001	33.20	31.220000	30.52
2	40800	776	33.24	47.520000	43.599998	21.84	22.100000	24.02
3	42900	393	45.72	58.720000	42.000000	34.10	36.239999	36.92
4	39600	610	41.70	59.220001	42.599998	29.86	29.400000	25.50
...	...	...	...	...	...	...	...	...
187957	81900	588	7.20	33.899999	66.400002	5.08	4.760000	2.74
187958	79800	732	14.74	33.519999	57.639999	9.56	11.420000	14.16
187959	79200	105	11.00	32.579999	41.599998	6.52	6.340000	6.50
187960	81900	486	2.30	17.380000	40.000000	1.32	2.060000	3.36
187961	80700	199	20.54	51.280001	39.900002	14.94	13.940000	16.00

187962 rows × 8 columns

Figure 3. Dataset summary of machine utilization file

This is machine utilization table. It includes utilization of cpu, memory and disk. Memory usage is 11.5 mb.

2.3. Online services

Online services are described by two tables: Service instance event and Service instance usage.

2.3.1. Online instance event(container_event.csv)

This table includes only two type of events : Create and Remove

Each 'create' event records the finish of an online instance creation, and each 'remove' event records the finish of an online instance removal. Online instance is given a unique cpuset allocation by online scheduler according to cpu topology and service constraints[4].

In [18]: data

Out[18]:

<table>
<tr><th></th><th>ts: timestamp of event</th><th>event: event type includes: Create and Remove</th><th>instance_id: online instance id</th><th>machine_id</th><th>plan_cpu: cpu number requested</th><th>plan_mem: normalized memory requested</th><th>plan_disk: normalized disk space requested</th><th>cpuset: assigned cpuset by online scheduler</th></tr>
<tr><td>0</td><td>0</td><td>Create</td><td>1000</td><td>1295</td><td>8</td><td>0.084819</td><td>0.056809</td><td>56|57|58|59|60|61|62|63</td></tr>
<tr><td>1</td><td>0</td><td>Create</td><td>10001</td><td>668</td><td>8</td><td>0.084819</td><td>0.056809</td><td>24|25|26|27|28|29|30|31</td></tr>
<tr><td>2</td><td>0</td><td>Create</td><td>10002</td><td>1217</td><td>8</td><td>0.084819</td><td>0.056809</td><td>4|5|6|7|8|9|10|11</td></tr>
<tr><td>3</td><td>0</td><td>Create</td><td>10003</td><td>1019</td><td>4</td><td>0.042409</td><td>0.034085</td><td>36|37|38|39</td></tr>
<tr><td>4</td><td>0</td><td>Create</td><td>10004</td><td>872</td><td>4</td><td>0.042409</td><td>0.034085</td><td>36|37|38|39</td></tr>
<tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>11096</td><td>39374</td><td>Create</td><td>4597</td><td>916</td><td>4</td><td>0.042409</td><td>0.034085</td><td>52|53|54|55</td></tr>
<tr><td>11097</td><td>39495</td><td>Create</td><td>6624</td><td>921</td><td>4</td><td>0.042409</td><td>0.034085</td><td>52|53|54|55</td></tr>
<tr><td>11098</td><td>42454</td><td>Create</td><td>10729</td><td>1096</td><td>4</td><td>0.042409</td><td>0.056809</td><td>56|57|58|59</td></tr>
<tr><td>11099</td><td>42454</td><td>Create</td><td>3101</td><td>1096</td><td>4</td><td>0.042409</td><td>0.056809</td><td>60|61|62|63</td></tr>
<tr><td>11100</td><td>79080</td><td>Create</td><td>1758</td><td>66</td><td>8</td><td>0.169637</td><td>0.056809</td><td>0|1|2|3|4|5|6|7</td></tr>
</table>

11101 rows × 8 columns

Figure 4. Dataset summary of online instance event file

There are total 11101 rows and 8 columns in the service instance event table. 11100 online requests come and All the online event is created.

2.3.2. Online instance usage(container_usage.csv)

Service instance usage is the total usage of cpu, memory and disk. This table has 12 columns as shown in fig 2.1

In [5]: data

Out[5]:

	ts: start time of measurement interval	instance_id: online instance id	cpu_util: used percent of requested cpus	mem_util: used percent of requested memory	disk_util: used percent of requested disk space	load1	load5	load15	avg_cpi, average cycles per instructions	avg_mpki: average last-level cache misses per 1000 instructions	max_cpi: maximum CPI	max_mpki: maximum MPKI
0	42600	107	3.30	24.000000	5.200000	0.54	0.38	0.30	0.155430	0.550153	2.211467	12.187318
1	42300	108	3.14	25.600000	10.600000	0.08	0.14	0.20	0.199342	0.294852	2.633724	3.997216
2	42000	109	3.82	42.000000	13.900000	0.10	0.16	0.20	0.238470	0.292426	2.203077	2.942878
3	41700	110	5.82	24.900000	7.400000	0.74	0.62	0.60	0.136100	0.161496	1.622428	2.469633
4	41400	111	3.92	24.000000	5.300000	0.40	0.32	0.28	0.200570	0.338131	2.758059	5.559884
...	...	...	...	...	...	...	...	...	...	...	...	...
1480900	80400	10175	4.32	34.440001	13.840000	0.12	0.18	0.20	0.174103	0.207152	2.024820	2.795278
1480901	80100	10176	3.72	27.180000	6.200000	0.18	0.14	0.10	0.133589	0.169097	1.657425	2.404944
1480902	79800	10177	14.90	52.060001	23.700001	0.76	0.76	0.76	0.085322	0.121745	1.207273	2.280462
1480903	79500	10178	3.20	27.700001	13.000000	0.20	0.20	0.20	0.154083	0.199628	2.690727	4.154834
1480904	79200	10179	25.48	64.639999	13.000000	2.58	2.66	2.98	0.169563	0.132053	1.620200	1.488545

1480905 rows × 12 columns

Figure 5. Dataset summary of online instance usage file

Data table has 1480905 rows and 12 columns. Total memory usage 135.6 mb.

2.4. Batch workload

Batch workload is described by two tables : instance table and task table.
Users submit batch workload in the form of Job (which is not included in the trace). A job contains multiple tasks, different tasks executes different computing logic.

2.4.1. Task table(batch_task.csv)

```
In [5]: data
Out[5]:
```

	create_timestamp	modify_timestamp	job_id	task_id	instance_num	status	plan_cpu	plan_mem
0	6457	6533	3	5	1	Terminated	50.0	0.004395
1	6036	6046	4	7	393	Waiting	NaN	NaN
2	6036	6046	4	6	452	Waiting	NaN	NaN
3	10719	11332	15	67	1705	Terminated	50.0	0.005736
4	10718	11164	15	66	631	Terminated	50.0	0.016007
...	...	...	...	...	...	...	...	...
80547	32996	33043	12935	80454	65	Terminated	50.0	0.009681
80548	32996	33061	12935	80453	69	Terminated	50.0	0.010706
80549	32996	33069	12935	80452	249	Terminated	50.0	0.007962
80550	32996	32999	12935	80457	1	Terminated	50.0	0.004059
80551	32996	33068	12935	80456	131	Terminated	50.0	0.007998

80552 rows × 8 columns

Figure 6. Dataset summary of task table file

task table is describe the task that is event status like Ready, Waiting, Running, Terminated, Failed, Canceled. Table has 80552 rows and 8 columns. Total memory usage 4.9+ mb.

2.4.2. Instance table(batch_instance.csv)

If machine fails or network problem than batch instance may fail. Each record in instance table record one try run. The start and end timestamp can be 0 for some instance status.

```
In [5]: data
Out[5]:
```

	start_timestamp: instance start time if the instance is started	end_timestamp: instance end time if the instance ended	job_id	task_id	machineID: the host machine running the instance	status	seq_no: running trials number	total_seq_no: total number of retries	real_cpu_max: maximum cpu numbers of actual instance running	real_cpu_avg: average cpu numbers of actual instance running	real_mem_max: maximum normalized memory of actual instance running	real_mem_avg: average normalized memory of actual instance running
0	41561	41619	120.0	686.0	1279.0	Terminated	1	1	0.89	0.28	NaN	NaN
1	41562	41617	120.0	686.0	828.0	Terminated	1	1	0.94	0.29	NaN	NaN
2	41561	41617	120.0	686.0	95.0	Terminated	1	1	1.00	0.31	NaN	NaN
3	41557	41610	120.0	686.0	545.0	Terminated	1	1	1.37	0.29	NaN	NaN
4	41557	41614	120.0	686.0	258.0	Terminated	1	1	1.18	0.27	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
16094650	73512	73514	NaN	NaN	365.0	Terminated	1	1	1.00	0.93	0.000813	0.000611
16094651	73512	73514	NaN	NaN	13.0	Terminated	1	1	1.00	0.35	0.001098	0.000906
16094652	73512	73518	NaN	NaN	421.0	Terminated	1	1	0.95	0.22	0.001512	0.000890
16094653	73512	73518	NaN	NaN	409.0	Terminated	1	1	0.93	0.25	0.001439	0.000947
16094654	73512	73519	NaN	NaN	815.0	Terminated	1	1	0.95	0.36	0.022245	0.021505

16094655 rows × 12 columns

Figure 7. Dataset summary of instance table file

III. ANALYSIS OF DATASET

3.1. Outliers of data

An outliers is an observation that lies an abnormal distance from other values in a random sample from a data. This definition leaves it up to the analyst to decide what will be considered abnormal. Before abnormal observations can be singled out, it is necessary to characterize normal observation.

Below is the box plot of outliers of each column of machine event table(server_event.csv).

```
In [7]: import seaborn as sns
        sns.boxplot(x=data['timestamp'])

Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x1fe42d0c7c0>
```

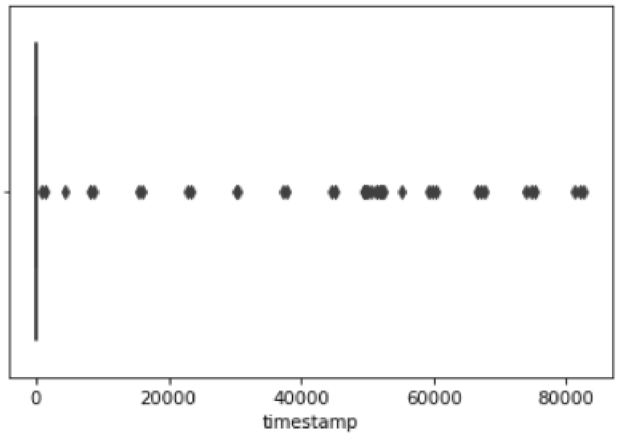


Figure 8. Outliers of timestamp column

In this Figure 8. all the dots are outliers means most of the data are same describe by line so allocate the machine to the process and outliers describes that soft error or hard error in this fig which means requests not get the machine.

```
In [8]: sns.boxplot(x=data['machineID'])

Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1fe4342ea90>
```

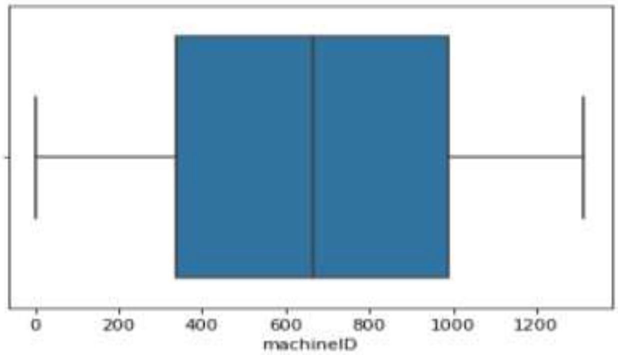


Figure 9. Outliers of machineID column

Figure 9 shows box plot of machineID column here is no dots means no outliers, most of machine ids are describes between 300 to 1000.

```
In [9]: sns.boxplot(x=data['capacity:CPU'])

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x1fe434c9ca0>
```

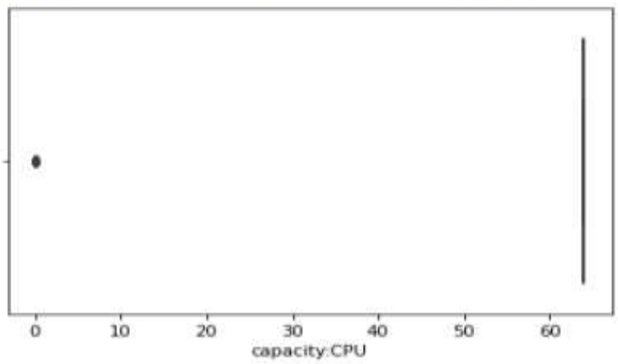


Figure 10. Outliers of capacity:CPU column

Figure 10. shows line that is capacity of cpu is same and only one dot means only one data from column is different.

```
In [10]: sns.boxplot(x=data['capacity:memory'])
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x1fe43529f40>
```

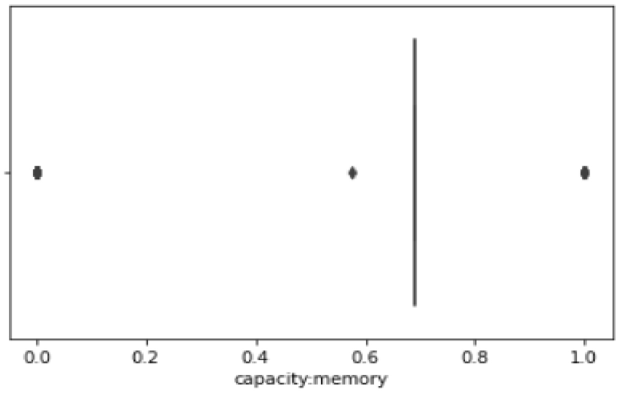


Figure 11. Outliers of capacity:memory column

Figure 11. shows outliers of capacity:memory column. It is same as capacity:CPU line describes that most of data is same and 2 or 3 values is different so this 2 or 3 values are outliers.

```
In [11]: sns.boxplot(x=data['capacity:disk'])
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1fe4357e4f0>
```

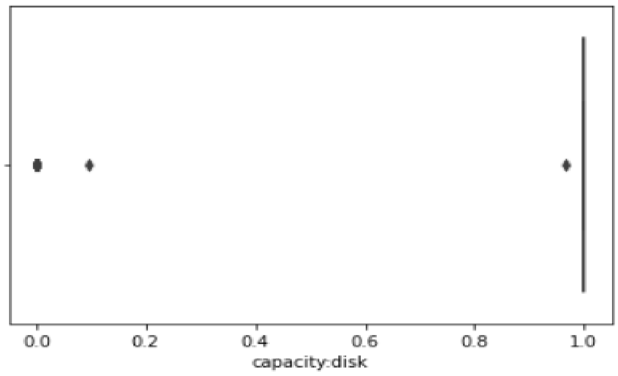


Figure 12. Outliers of capacity:disk column

Figure 12. describe capacity of disk so the capacity of disk is 1.

3.2. Utilization

There is 3 types of utilization in the alibaba dataset : CPU utilization, Memory utilization and Disk utilization

3.2.1. CPU utilization

$$\text{CPU utilization} = \frac{\text{sum of avg use of CPU in each amchine}}{1300}$$

$$\text{Avg use of CPU} = \frac{\text{Sum of used CPU per machine}}{144}$$

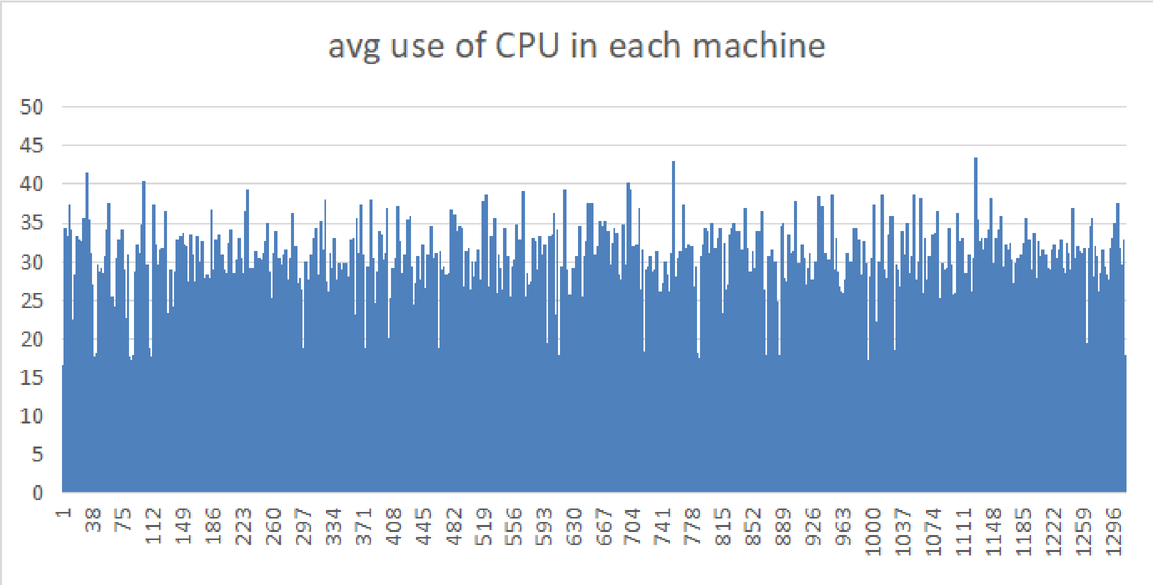


Figure 13. Utilization of CPU

Figure 13 represents the average use of CPU in each machine. Here all CPU used under 50% in each machine. In figure row represents the 1300 machines and column represent used of CPU in percentage. Average CPU utilization is 26.572% is shown in Figure 14.

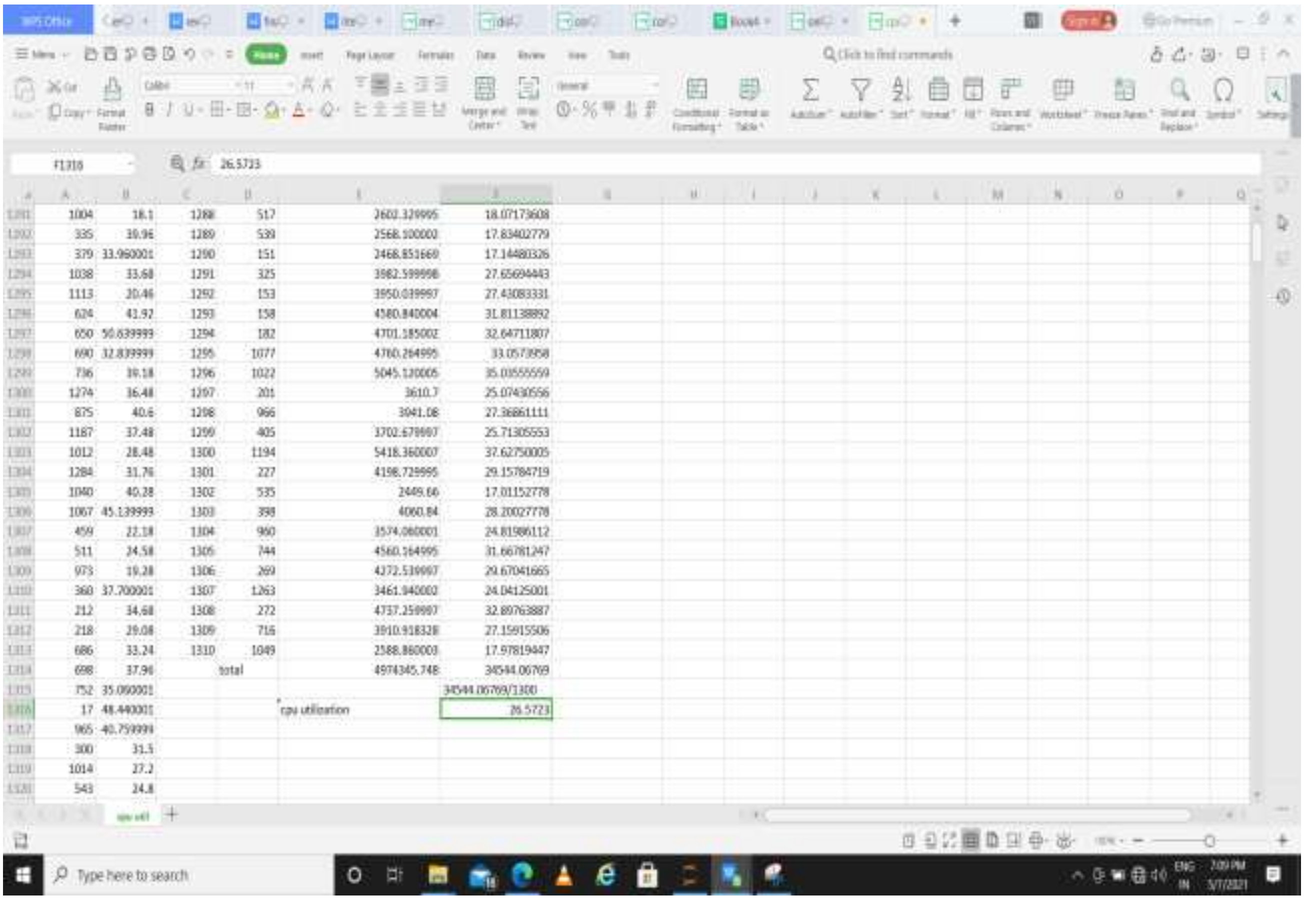


Figure 14. CPU utilization in percentage

3.2.2. Memory utilization

$$\text{Memory utilization} = \frac{\text{Sum of avg use of memory in each amchine}}{1300}$$

$$\text{Avg use of memory} = \frac{\text{Sum of used memory per machine}}{144}$$

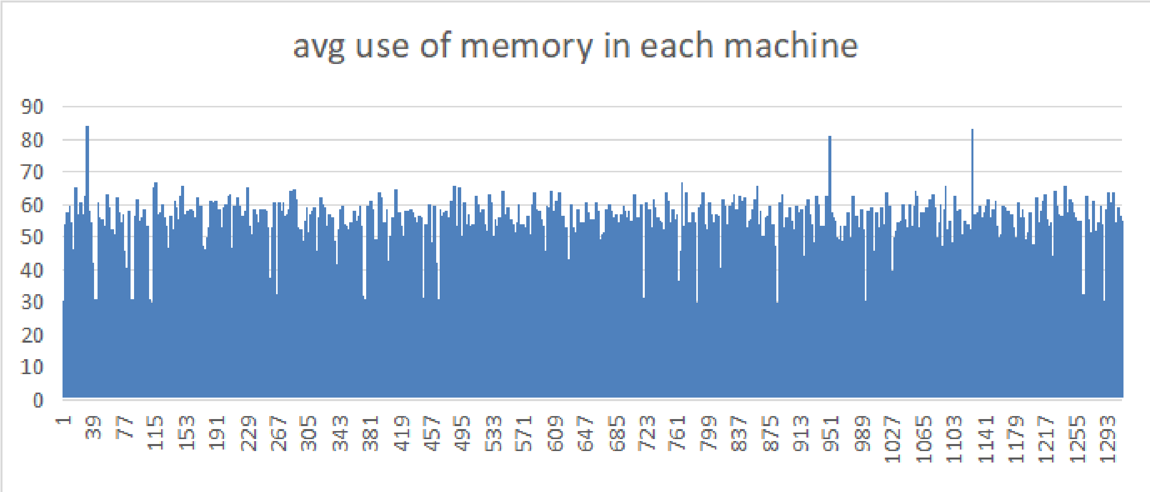
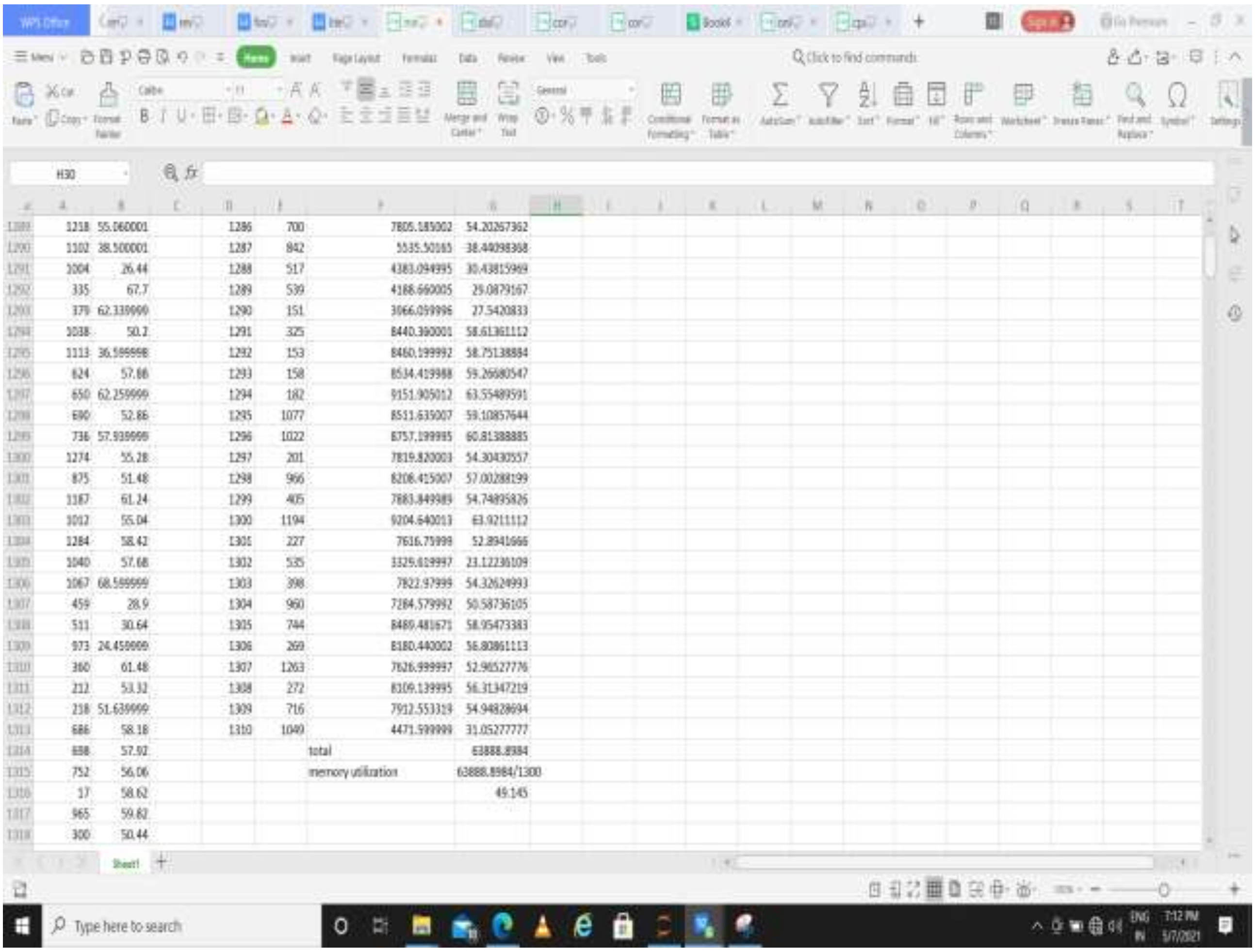


Figure 15. Utilization of memory

Figure 15. represents the usage of memory in each machine. Here overall memory usage is under 70% only two or three machine's memory usage are up to 80%. so the memory utilization is 49.145% as shown in Figure 16.



	A	B	C	D	E	F	G
1289	1218	55.060001		1286	700	7805.185002	54.20267362
1290	1102	38.500001		1287	842	5535.50165	38.44098368
1291	1004	26.44		1288	517	4383.094995	30.43815969
1292	335	67.7		1289	539	4188.660005	29.0879167
1293	379	62.339999		1290	151	3966.059996	27.5420833
1294	1038	50.2		1291	325	8440.390001	58.61361112
1295	1113	36.599998		1292	153	8460.199992	58.75138884
1296	624	57.88		1293	158	8534.419988	59.26680547
1297	650	62.259999		1294	182	9151.905012	63.55489591
1298	690	52.86		1295	1077	8511.635007	59.10857644
1299	736	57.939999		1296	1022	8757.199985	60.81388885
1300	1274	55.28		1297	201	7819.820003	54.30430557
1301	875	51.48		1298	966	8208.415007	57.00288199
1302	1187	61.24		1299	405	7883.849989	54.74895826
1303	1012	55.04		1300	1194	9204.640013	63.9211112
1304	1284	58.42		1301	227	7616.75999	52.8941666
1305	1040	57.68		1302	535	3329.619997	23.12236109
1306	1067	68.599999		1303	398	7822.97999	54.32624993
1307	459	28.9		1304	960	7284.579992	50.58736105
1308	511	30.64		1305	744	8489.481671	58.95473383
1309	973	24.459999		1306	269	8180.440002	56.80861113
1310	360	61.48		1307	1263	7626.999997	52.96527776
1311	212	53.32		1308	272	8109.139995	56.31347219
1312	218	51.639999		1309	716	7912.553319	54.94828694
1313	686	58.18		1310	1049	4471.599999	31.05277777
1314	698	57.92				total	63888.8984
1315	752	56.06				memory utilization	63888.8984/1300
1316	17	58.62					49.145
1317	965	59.82					
1318	300	50.44					

Figure 16. Memory utilization in percentage

3.2.3. Disk utilization

$$\text{disk utilization} = \frac{\text{sum of avg use of disk in each amchine}}{1300}$$

$$\text{Avg use of disk} = \frac{\text{Sum of used disk per machine}}{144}$$

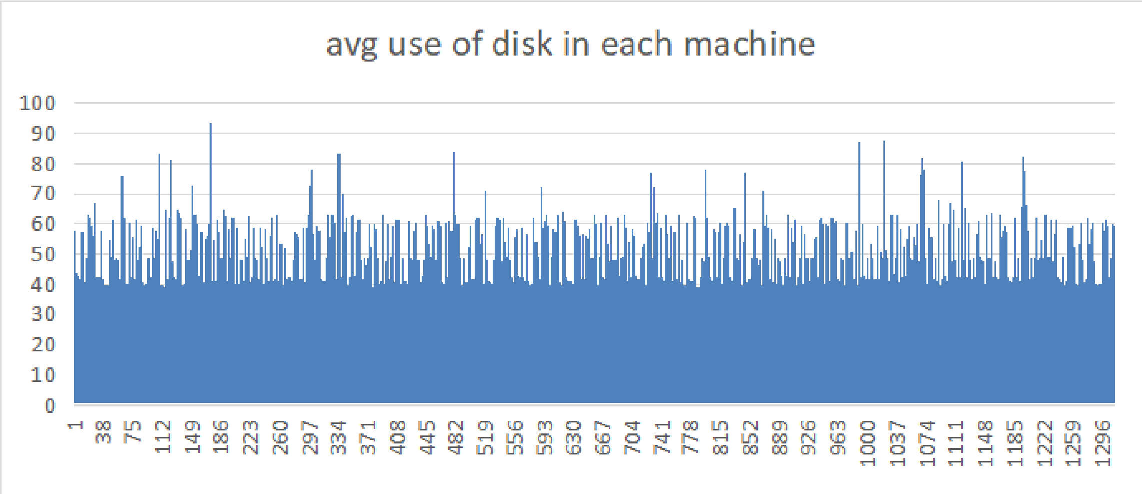
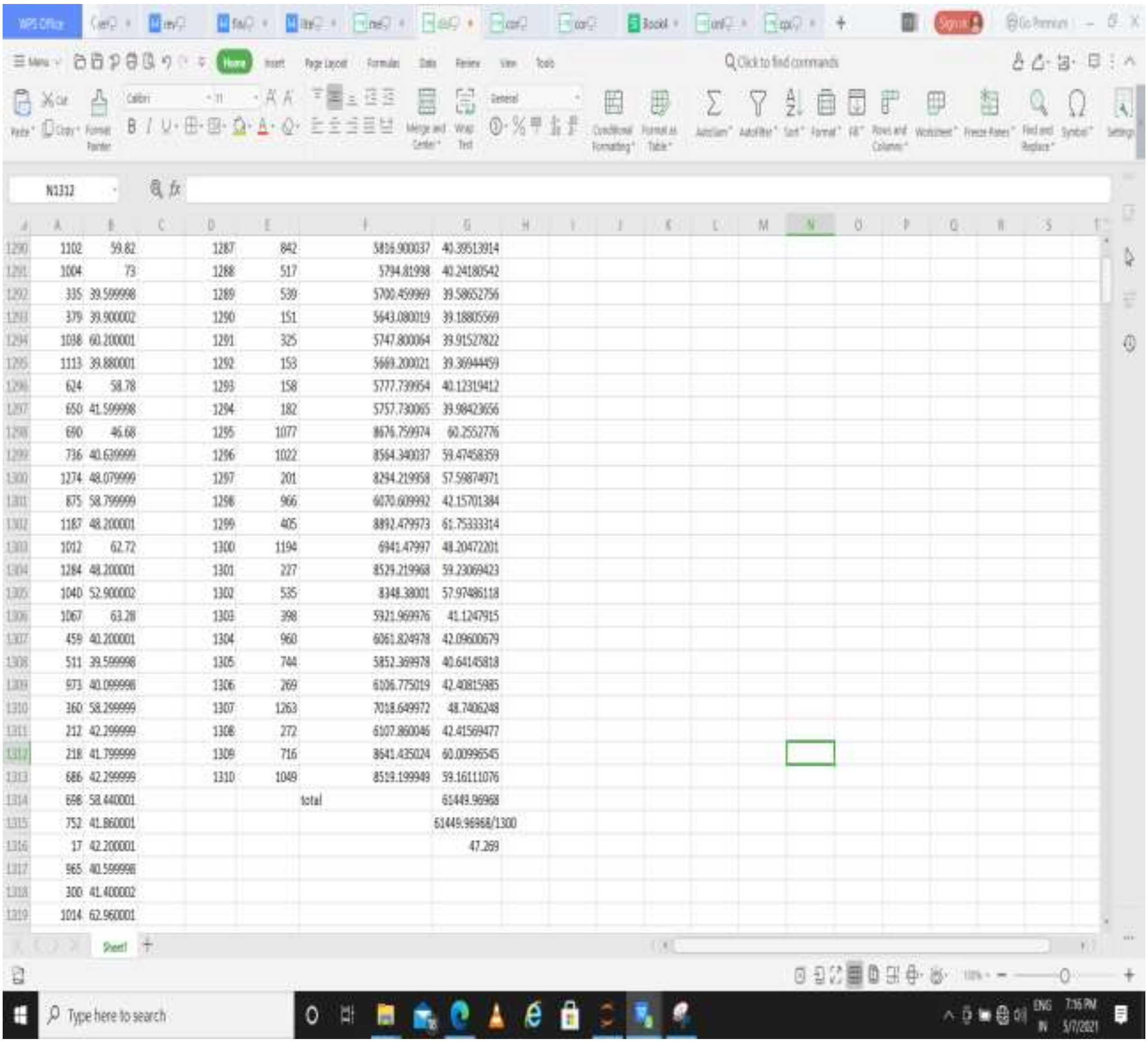


Figure 17. Utilization of disk

Figure 17. represents the usage of disk in each machines. Here half of disk are used up to 40% and half of disk used up to 70%. average usage of disk is 47.26% as shown in Figure 18.



	A	B	C	D	E	F	G
1290	1102	59.82		1287	842	5816.900037	40.39513914
1291	1004	73		1288	517	5794.81998	40.24180542
1292	335	39.599998		1289	539	5700.459969	39.58652756
1293	379	39.900002		1290	151	5643.080019	39.18805569
1294	1038	60.200001		1291	325	5747.800064	39.91527822
1295	1113	39.880001		1292	153	5669.200021	39.36944459
1296	624	58.78		1293	158	5777.739954	40.12319412
1297	650	41.599998		1294	182	5757.730065	39.98423656
1298	690	46.68		1295	1077	8676.759974	60.2552776
1299	736	40.639999		1296	1022	8564.340037	59.47458359
1300	1274	48.079999		1297	201	8294.219958	57.59874971
1301	875	58.799999		1298	966	6070.609992	42.15701384
1302	1187	48.200001		1299	405	8892.479973	61.75333314
1303	1012	62.72		1300	1194	6941.47997	48.20472201
1304	1284	48.200001		1301	227	8529.219968	59.23069423
1305	1040	52.900002		1302	535	8348.38001	57.97486118
1306	1067	63.28		1303	398	5921.969976	41.1247915
1307	459	40.200001		1304	960	6061.824978	42.09600679
1308	511	39.599998		1305	744	5852.369978	40.64145818
1309	973	40.099998		1306	269	6106.775019	42.40815985
1310	360	58.299999		1307	1263	7018.649972	48.7406248
1311	212	42.299999		1308	272	6107.860046	42.41569477
1312	218	41.799999		1309	716	8641.435024	60.00996545
1313	686	42.299999		1310	1049	8519.199949	59.16111076
1314	698	58.440001				total	61449.96968
1315	752	41.860001					61449.96968/1300
1316	17	42.200001					47.269
1317	965	40.599998					
1318	300	41.400002					
1319	1014	62.960001					

Figure 18. Utilization of disk in percentage

IV. EXISTING WORK ON CURRENT DATASET

According to the authors: Yue Cheng, Zheng Chai, Ali Anwar, In 2011, Google open-sourced the first publicly available cluster trace data spanning several clusters[2]. Reiss et al. study the differently and dynamicity properties of the Google workloads. Alibaba, the largest cloud service provider in China, released their cluster trace [1] in late 2017. Different from the Google trace, the Alibaba trace contains information about the two located container and batch job workloads, facilitating better understanding of their interactions and interferences. Lu et al. perform characterization of the Alibaba trace to reveal basic workload statistics. Their study is focused on providing a unique and microscopic view about how the co-located workloads interact and impact each other[11].

According to the authors: CONGFENG JIANG 1,2, (Member, IEEE), GUANGJIE HAN 3 , (Member, IEEE), JIANGBIN LIN4 , GANGYONG JIA1,2, (Member, IEEE), WEISONG SHI5 , (Fellow, IEEE), AND JIAN WAN1,6(Member, IEEE), To explore various characteristics of allocated online services and batch jobs from a production cluster containing 1300 servers in Alibaba Cloud. From the trace data, we find the following: 1) For batch jobs with multiple tasks and instances, 50.8% failed tasks wait and halted after a very long time interval when their first and the only one instance fails. 2) For online services jobs, they are clustered in 25 categories according to their requested CPU resources, memory resources, and disk resources. Such clustering can help the co-allocation of online services jobs with batch jobs. 3) Servers are clustered into seven groups by CPU and memory utilization and their correlations. Machines with a strong correlation between CPU and memory utilization provides an opportunity for job co-allocation and resource utilization estimation. 4) They also compare the cumulative distribution functions of jobs and servers and explain the differences and opportunities for workload assignment between them[10].

Authors:Chengzhi Lu1 , Kejiang Ye1 , Guoyao Xu1,2 , Cheng-Zhong Xu1 , Tongxin Bai1 said that, to improve resource efficiency and design intelligent scheduler for clouds, it is necessary to understand the workload characteristics and machine utilization in high-scale cloud data centers.Analysis reveals several important insights about different types of imbalance in the Alibaba cloud. Such imbalances aggravate the complexity and challenge of cloud resource management, which might incur severe wastes of resources and low cluster utilization. 1) Spatial Imbalance: different resource utilization across machines and workloads. 2) Temporal Imbalance: time-varying resource usages per workload and machine. 3) Imbalanced proportion of multi-dimensional resources (CPU and memory) utilization per workload. 4) Imbalanced resource demands and runtime statistics between online service and offline batch jobs. We argue accommodating such imbalances during resource allocation is critical to improve cluster efficiency, and will motivate the emergence of new resource managers and schedulers[1].

According to the authors: Yizhou Shan, Yutong Huang, Yilun Chen, Yiying Zhang, to improve resource utilization, elasticity, heterogeneity, and failure handling in datacenters, they believe that datacenters should break monolithic servers into disaggregated, network-attached hardware components. Despite the promising benefits of hardware resource disaggregation, no existing OSes or software systems can properly manage it[7].

CONCLUSION

This is a review of the ongoing research and analysis done in this field of Alibaba cloud dataset. This paper performs fundamental analysis like datatype, minimum value, maximum value, outliers of data, CPU utilization, memory utilization, and disk utilization by the formula written in this paper. From the analysis we concluded that utilization of CPU is 26.57%, utilization of memory is 49.14% and the utilization of the disk is 47.26%. At last, discussed various techniques that can apply to this dataset for better resource utilization.

REFERENCES

[1] Author name : Chengzhi Lu1, Kejiang Ye1, Guoyao Xu1, Cheng-Zhong Xu1, Tongxin Bai1, Imbalance in the Cloud: an Analysis on Alibaba Cluster Trace; publication year : 2017 , IEEE International Conference on Big Data.
[2] A. Verma, L. Pedrosa, M. Korupolu, D. Oppenheimer, E. Tune, and J. Wilkes, “Large-scale cluster management at google with borg,” in Proceedings of the Tenth European Conference on Computer Systems. ACM, 2015, p. 18.
[3] clusterdata/cluster-trace-v2017,https://github.com/alibaba/clusterdata/blob/master/cluster-trace-v2017/trace_201708.md
[4] clusterdata/cluster-trace-v2017, https://github.com/alibaba/clusterdata
[5] B. Hindman, A. Konwinski, M. Zaharia, A. Ghodsi, A. D. Joseph, R. H. Katz, S. Shenker, and I. Stoica, “Mesos: A platform for fine-grained resource sharing in the data center.” in NSDI, vol. 11, 2011, pp. 22–22.
[6] M. Schwarzkopf, A. Konwinski, M. Abd-El-Malek, and J. Wilkes, “Omega: flexible, scalable schedulers for large compute clusters,” in Proceedings of the 8th ACM European Conference on Computer Systems. ACM, 2013, pp. 351–364.

[7] Authors name:Yizhou Shan, Yutong Huang, Yilun Chen, Yiying Zhang LegoOS: A Disseminated, Distributed OS for Hardware Resource DisaggregationOct-18, 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI '18)

[8] E. Boutin, J. Ekanayake, W. Lin, B. Shi, J. Zhou, Z. Qian, M. Wu, and L. Zhou, "Apollo: Scalable and coordinated scheduling for cloud-scale computing," in Proceedings of the 11th USENIX Conference on Operating Systems Design and Implementation, ser. OSDI '14. USENIX Association, 2014, pp. 285–300.

[9] Z. Zhang, C. Li, Y. Tao, R. Yang, H. Tang, and J. Xu, "Fuxi: a fault-tolerant resource management and job scheduling system at internet scale," in Proceedings of the VLDB Endowment, ser. VLDB '14. VLDB Endowment, 2014, pp. 1393–1404.

[10] Author name: CONGFENG JIANG 1, (Member, IEEE), GUANGJIE HAN 3,(Member, IEEE), JIANGBIN LIN4,GANGYONG JIA1,2, (Member, IEEE), WEISONG SHI5, (Fellow, IEEE),AND JIAN WAN1,6(Member, IEEE) Characteristics of Co-Allocated Online Services and Batch Jobs in Internet Data Centers: A Case Study From Alibaba Cloud publication year: 2019, IEEE Access

[11] Authors name: Yue Cheng, Zheng Chai, Ali Anwar, Characterizing Co-located Datacenter Workloads: An Alibaba Case Study publication year : 2018

A novel approach to workload prediction using attention-based LSTM encoderdecoder network in cloud environment Yonghua Zhu1,2, Weilin Zhang1 , Yihai Chen1,3 and Honghao Gao4,5*, springer publication, publication year : Dec 2019.

